ASAVAL Conference

Valuation for the Financial Sector:

Challenges in the short and medium term

19th November, 2021

AVMs – Opportunity or Threat?

A holistic view of technology at the service of the evaluator

Professor George Matysiak

Agenda

- What are Automated Valuation models (AVMs)?
- AVMs and appraisals
- AVM valuation accuracy
- Evaluating AVMs
- AVMs and data
- Appendix: Regression-based AVM models

Automated Valuation Models (AVMs)

- AVMs have their origins in North America, the first commercial application being in the 1980s. AVMs began to be developed in the UK in the 1990s. The widespread use of AVMs is now firmly established in many counties.
- Although traditional approaches are routinely employed in the valuation profession, there has been a significant growth in independent residential Automated Valuation Model (AVM) providers, who offer their services routinely on a fee-based basis, to both lenders and the fee-paying public.
- AVMs are widely used by lenders and institutional investors, largely for monitoring purposes, and are seen as complementary to traditional valuations.
- In the EU, AVMs will be allowed to be used for loan origination (*Capital Requirement Regulations*).
- These *computer-assisted quantitative methods* have advantages in that they are systematic and fast, thereby reducing reliance on labour input in providing an end-to-end valuation.
- By removing the human element, it is claimed by some advocates, it also reduces inaccuracies due to reliance on human judgement.
- The overall attitude and degree of acceptance of such automated approaches to valuation varies.
- AVMs are now viewed in the wider context of the advance of *PropTech (digital approaches)* within the real estate industry.

What is an AVM 1?

- Over the last number of years, property valuation has evolved from manual sales comparison methods, and subjective valuer assessments based on simply comparing properties, into *mechanical* oriented valuation models.
- Although different underlying AVM models are employed by vendors, fundamental to the approach are *statistical, data mining and computing technicalities.*
- TEGoVA provide the following, Definition 2.1, in their European Valuation Standards EVIP 6:

    'Automated Valuation Models (AVMs) can be defined as statistic-based computer programmes, which use property information (e.g. comparable sales and property characteristics etc.) to generate property-related values or suggested values.'

- The International Association of Assessing Officers, IAAO (2003), describes an AVM as:

'a mathematically based computer software programme that produces an estimate of market value based on analysis of location, market conditions, and real estate characteristics from information collected. The distinguishing feature of an AVM is that it produces a market valuation through mathematical modelling. The credibility of an AVM is dependent on the data used and the skills of the modeller producing the AVM.'

What is an AVM 2?

- The following definition of an Automated Valuation Model is provided by the RICS AVM Standards Working Group:

  'Automated Valuation Models use one or more mathematical techniques to provide an estimate of value of a specified property at a specified date, accompanied by a measure of confidence in the accuracy of the result, without human intervention post-initiation.' (RICS 2013).

- A key component in the RICS definition is the qualification

  '…accompanied by a measure of confidence in the accuracy of the result…'.

- All three definitions of an AVM *exclude* any appraiser involvement in arriving at a value.

AVM models

- AVMs reflect ‘patterns’ in the behaviour of groups of properties.
- Multiple regression analysis (MRA) and hedonic models are extensively employed in estimating AVMs.
- However, over the last few years alternative approaches employing machine learning applications have increasingly been used in AVMs.
- The quality of AVMs will vary considerably, depending on data preparation, sample size and the design of the ‘model’.

Data considerations

- Models are only as good as the data that is used in estimating them:

  (garbage in = garbage out!)
- If the quality of data is suboptimal (i.e. missing or inaccurate), estimates will be inaccurate.
- Current short comings of AVMs is much more information (data) based rather than methodology or modelling related aspects.
- The current condition of many properties is largely unknown, given the typical property level of datasets being used. Assumed that the property is in 'average' condition.
- Data quality improvements at the micro-data level are required. The technologies are there but the costs currently inhibit across the board adoption.
- As the data access becomes cheaper and more broadly available, the data will be incorporated into automatic valuation models, with the expectation that 'error' rates will continue to decline.

Acknowledgement: Insight and observations credited to discussions with HouseCanary.

Practitioner attitudes towards AVMs

- An international survey undertaken in 2008 on AVMs and the integration of AVMs within the valuation process provides some interesting findings. There were 473 valuer responses, representing both lending and valuation organisations, and described as senior professional members with ‘much experience of mortgage valuations’. The results of the survey include the following :
- 71% of the valuers agreed that AVMs were inadequate for loan valuations as a result of no physical inspection.
- 87% of the valuers agreed that physical valuations were more accurate than AVMs, as a result of local knowledge.
- 90% of valuers agreed that the ability to evaluate comparables was a major advantage over AVMs.
- Has the position changed over the last 13 years?

Information on AVMs

- Debate regarding the role and accuracy of AVM valuations is an ongoing topic of debate.
- Whilst there are a large number of AVM vendors, the inner functioning of the models and details of their specification are not disclosed.
- Vendors claim they test their models regularly for accuracy, and inceasingly have the figures independently assessed.
- Other than submitting information to rating agencies, European/UK AVM operators, for example, are unwilling to have their data/methodologies exposed to independent scrutiny.
- US AVM market is highly developed and ‘accuracy’ figures are available on websites.

Most (all?) AVM valuations will be ‘wrong’ Why?

- Model errors: the model may be mis-specified and/or does not adequately capture market pricing i.e. the model is not perfect.
- AVMs are unable to account for buyer and seller motivation!
- Model stability: structural breakdown/model failure – the world has changed!
- Markets are not static – market conditions change.
- Quality of data and outdated information.
- Lack of ‘relevant’ data i.e. model calibrated on data which is outside of range of the property being valued.
- Thin/few transactions in the market.
- Random component at work in all markets
- Always look at the prediction interval – FSD (range)!

Measuring an AVM’s accuracy

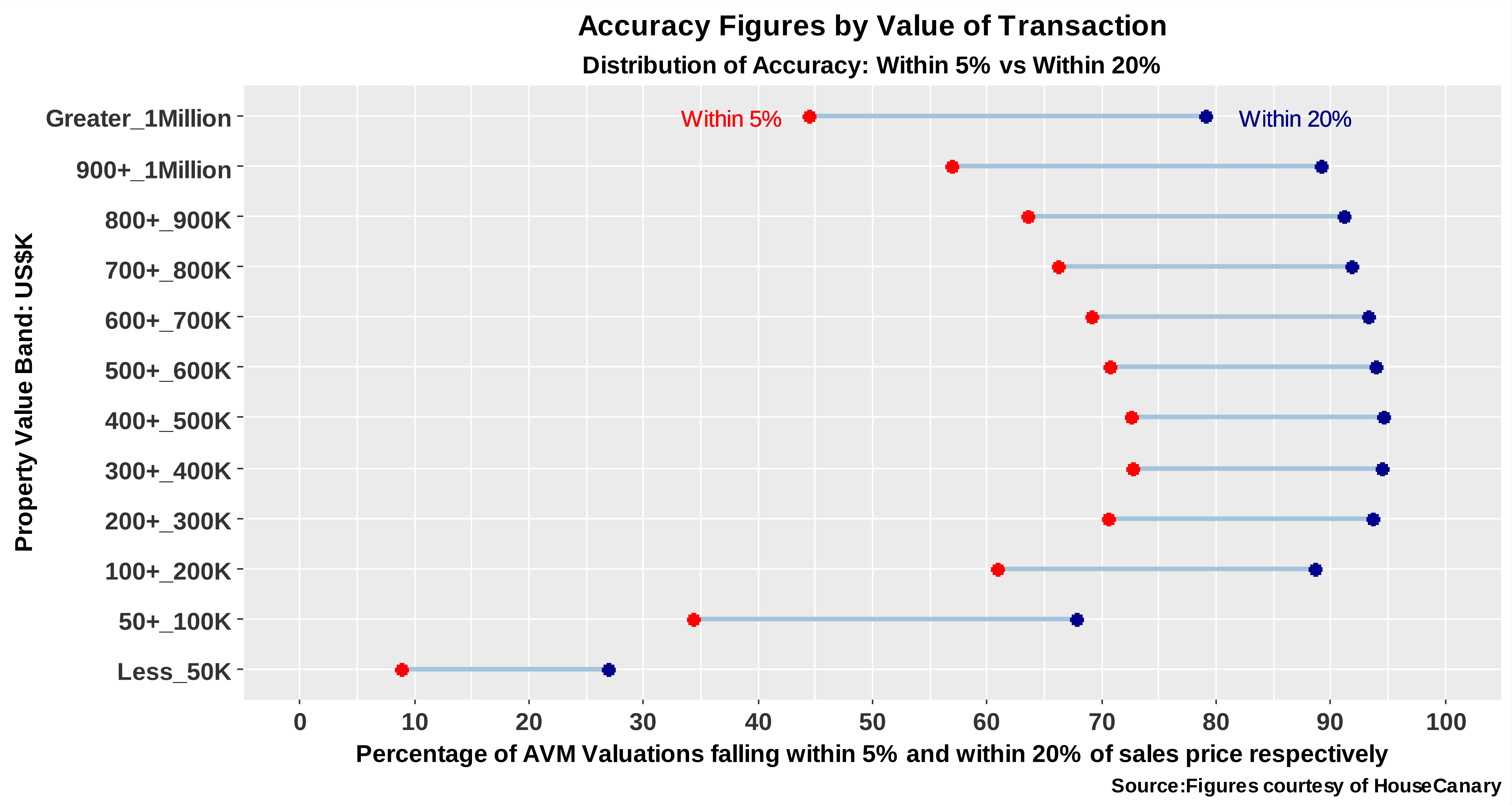
- There are a variety of ways to measure AVM valuation error (refer to the *TEGoVA* report). First, a *benchmark reference* needs to be established:

  - is the AVM forecast measured relative to a valuer’s estimate of the property’s price or,

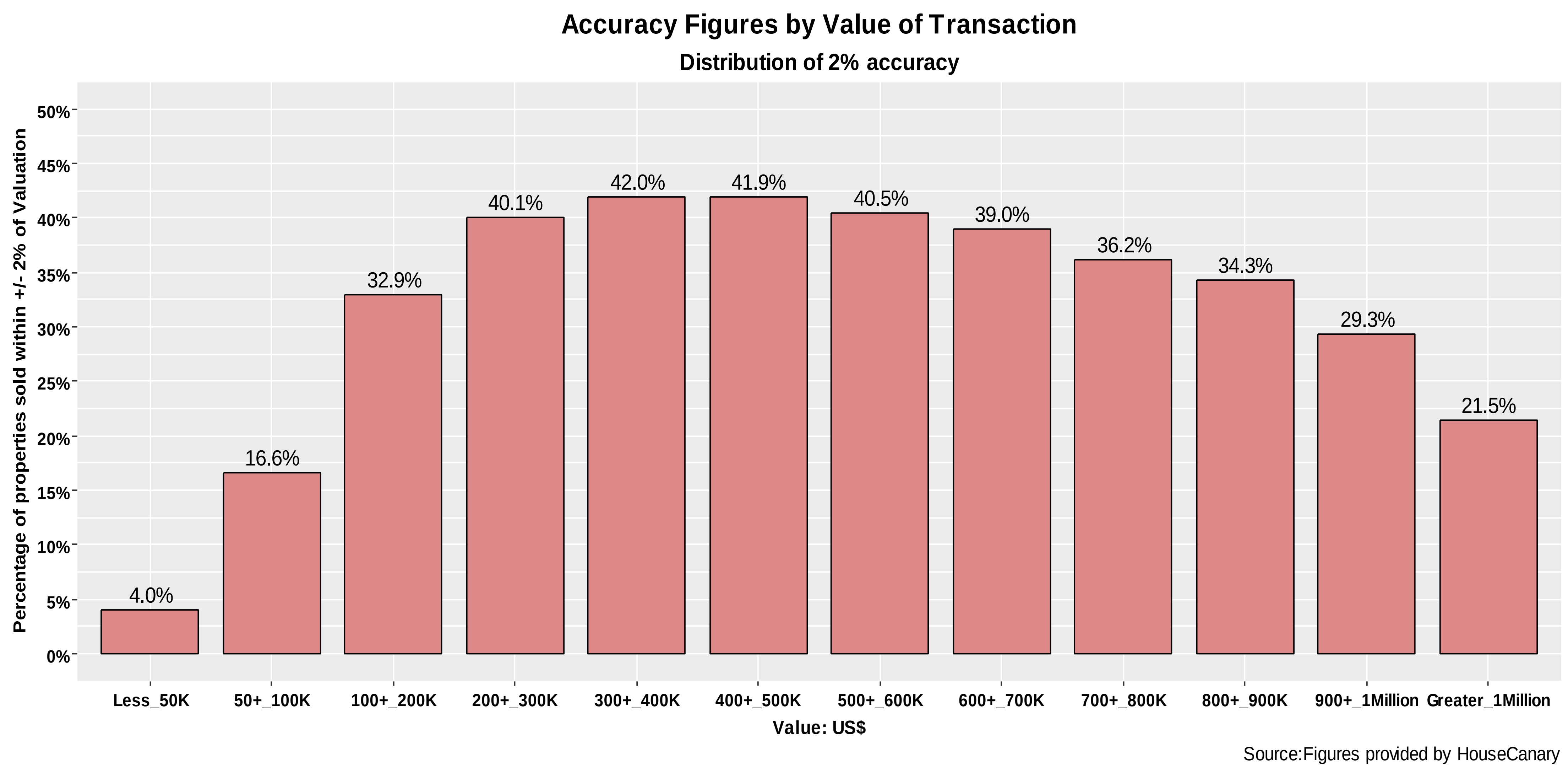
  - is the AVM forecast measured relative to the market price achieved in the market?

- Several US vendors and independent testers of AVMs (e.g. *AVMetrics*) publish accuracy results based on achieved sales prices.
- AVM vendors typically qualify their valuation estimates by providing a *prediction range* with a specified degree of confidence.

Distribution of valuation accuracy: HouseCanary



Distribution of valuation accuracy within +/-2% of sales price: HouseCanary



Valuation error – ‘noise’

- All models will contain error – you cannot 100% explain everything!
  - cannot model noise!
- There will always be random ‘errors’ in AVM valuations (indeed, in any financial/econometric model).
- AVM modellers need to recognise the presence of noise *and* quantify it.
  - beware of danger in over-fitting models!
- Laws and regulations (EU *General Data Protection Regulation – GDPR*) are cropping up to restrict the blind use of complex models. EU countries must adhere to a new regulation *requiring algorithmic model decisions to be explainable to citizens directly affected*.
  - will this regulation apply to AVMs?
  - if so, how does one explain a ‘black box’ AVM model?

    - considerable progress has been made on this front!

Explaining model generated results

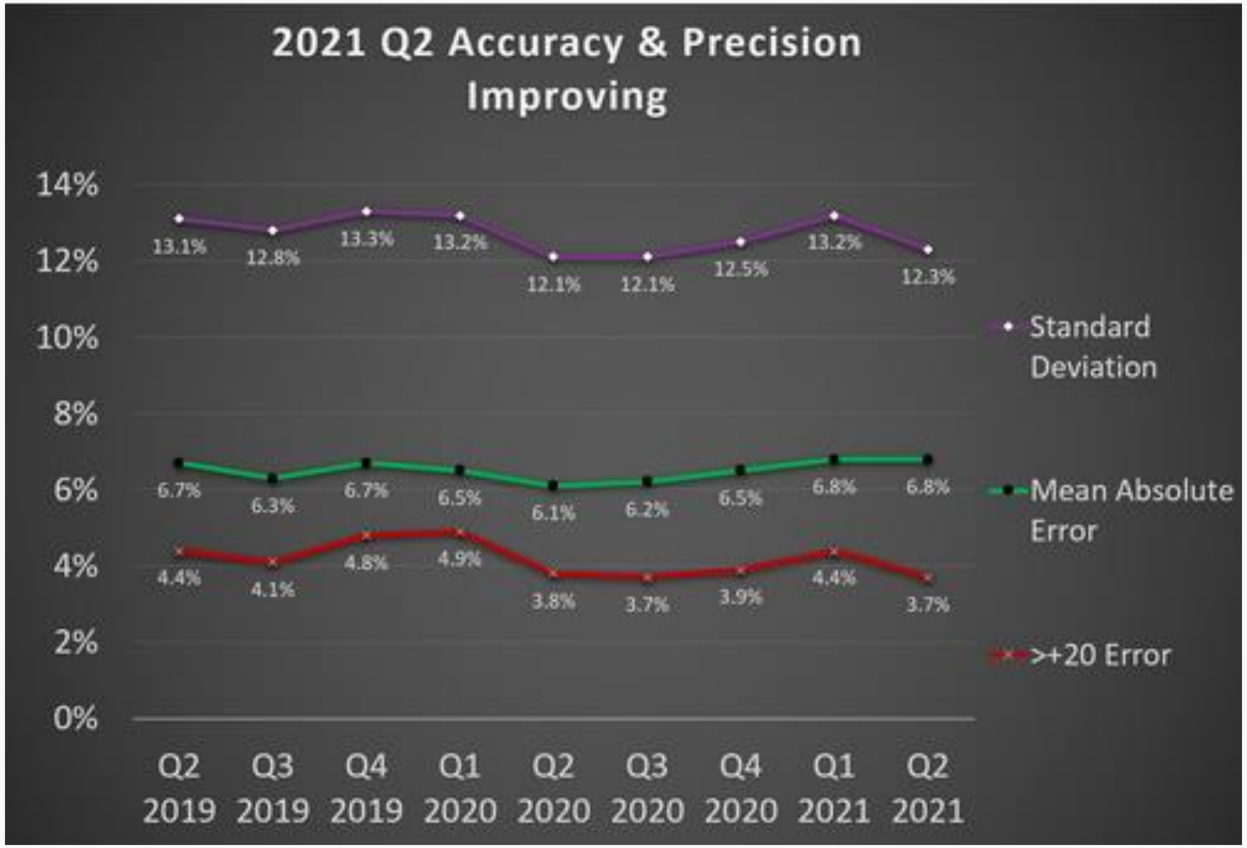
- Laws and regulations (EU *General Data Protection Regulation – GDPR*) are cropping up to restrict the blind use of complex models. All EU countries must adhere to a new regulation *requiring algorithmic decisions to be explainable to citizens directly affected*.
  - will this regulation apply to AVMs?
  - if so, techniques are available
  - …an AVM prediction?

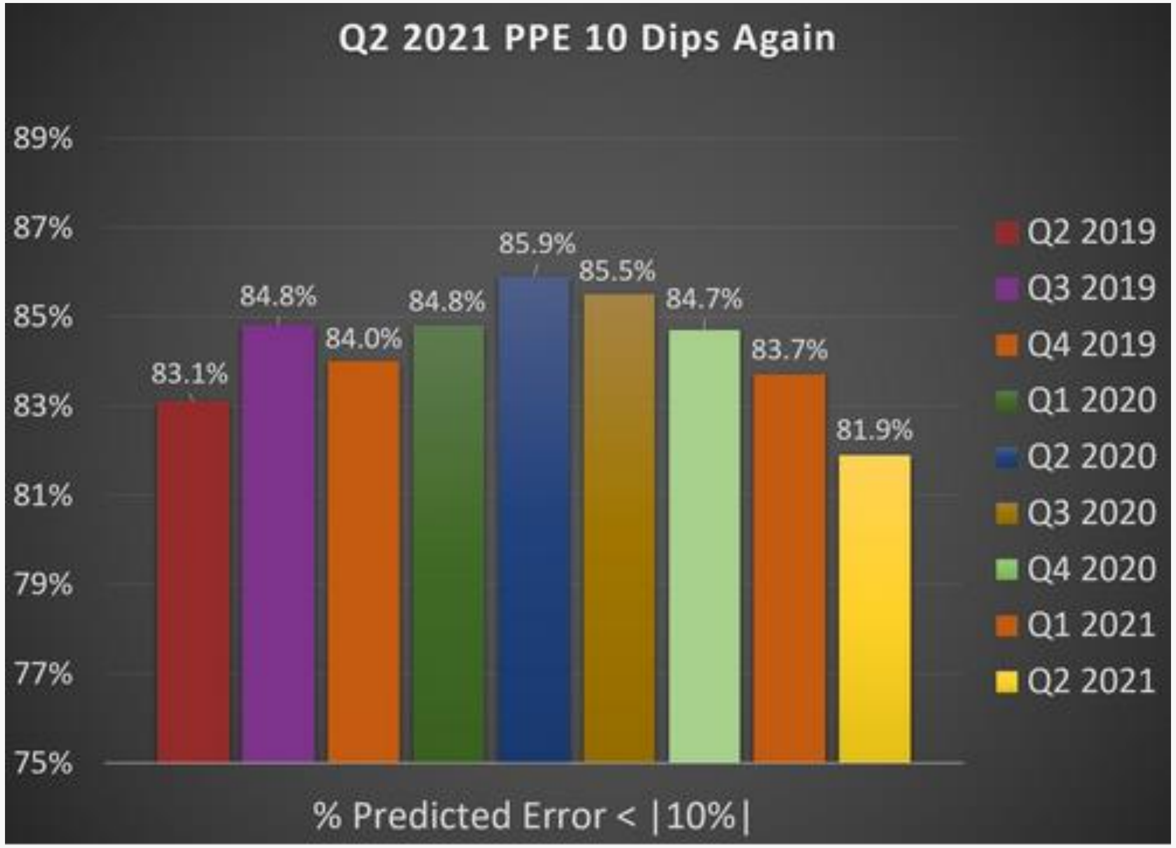
AVM valuation under different conditions

- Is the margin for error unchanging i.e. constant? It will most likely vary under different conditions.
- There are a whole set of circumstances which would need to be taken into consideration when looking to assess what would be an acceptable margin of error in valuing residential properties, including the following:
    - different market conditions/environments, such as rising/falling prices
    - up market *versus* down market (asymmetric effect)
    - different size/value properties
    - quality of property
    - age of property
    - market liquidity e.g. dependent on the volume of transactions
    - geographic location/different neighbourhoods
    - type of property
- All of the above are likely to vary *by* country and *within* country!
- Question: Can sample specific errors be assumed to be ‘typical’?
- Do AVM models pick-up on the variety of underlying factors driving prices?

AVM accuracy across US counties in 2021

2021 Q2 figs Based on 1,000,200 transactions in 2650 counties and 24 AVM models





Observation from AVMetrics: *“The best AVM changes all the time”*

Source: AVMetrics, Newsletter No 9, June 2021

A word of caution

AVMetrics research reported on AVM accuracy figures in the US:

“The AVM-by-FSD analysis reveals that AVM providers are empirically under-reporting their vendor-reported FSDs for 267 (72.8%) of the 367 AVM/FSD combinations, which gives the appearance that AVM valuations are more precise than they really are”.

** “AVM Testing and Evaluation using AVM Performance Metrics”, AVMetrics White paper,2021*

AVMs, data considerations and questions

- Differences in models *and* the accuracy of the data affects the AVM results
- Sources of information – is it reliable and is the data reliable
    - Is the data accurate
    - Is up-to-date-current
    - Is the data comprehensive/complete
    - Is the data regularly (daily?) updated or are there time-lags (issues when markets rising/falling rapidly)
- Are the transactions recent or dated
- Are the properties comparable (“comps”) and sold recently
- How many comps are used, are they really comps i.e. ‘similar’ – what if the market is thin with few transactions
- Data issues in non-transparent markets such as Portugal
- Can model adjust generated value take account of quality of a property
    - some on-line valuation tools allow user generated specific details about their property
    - online property value estimators (AVMs) very unlikely to provide the same value:

      ⇨ different models + different data
    - how large are the differences
- Are uncertainty/value range/error rate measures reported ⇨ ‘comfort factor’
- Should AVM estimates be validated by a valuer ⇨ ‘bigger comfort factor’

Data considerations: summary

- Data quality improvements at the micro-data level are required. The technologies are there but costs currently inhibit across the board adoption.
- Improvement in AVM models will result because of increased access to more finely grained data.
- A general assumption is that only marginal gains will result from new algorithms.

Acknowledgement: Insight and observations credited to discussions with HouseCanary.

A role for the valuer?

- AVMs are extensively used around the world and have become part of the valuation landscape. Given the widespread deployment of AVMs, their use is not an either-or question, but a question of *how* can an AVM enhance a valuer's estimate of value.
- The position has been well summarised by a significant provider of valuation technologies to the mortgage banking industry as follows:

> *"AVMs are going to get more and more mainstream, particularly as data and analytics get more sophisticated. AVMs won't take the place of an appraisal. There will always be a need for local knowledge and expertise, not to mention an on-site evaluation of the physical property."*

- TEGoVA has issued a European valuation standard and guidance note on AVMs:

European Valuation Standard 6

AVMs cannot be used to produce a valuation report that complies with EVS independently of a valuation process founded, inter alia, on inspection of the property by the valuer and the application of valuation judgment by the valuer.

Where used, an AVM is never more than a tool contributing to the valuer's estimation of value, for which he remains responsible.

Challenges in the short and medium term

- The independent validation and standards of validation of European AVMs needs to be promoted more vigorously, otherwise the role of AVMs will continue to be contested. How best to proceed?

- In the absence of regulatory/enforceable controls, there has to be a commercial or reputational benefit to the vendor in order to make it worthwhile for them to provide more model based information.

- Assessing AVM 'quality':
  - Independent professional bodies qualified to scrutinise AVMs
  - setting of standards of best practice for AVM vendors
  - access to the underlying models
  - access to database(s) on which the models are calibrated/estimated/tested
  - access to AVM output under different market environments and to any 'adjustment' made to pure model generated forecasts
  - AVM accuracy/standardisation of published accuracy measures
  - clear definitions/standards for transparent 'testing' of AVMs procedures

ASSESSING THE ACCURACY OF INDIVIDUAL PROPERTY VALUES ESTIMATED BY AUTOMATED VALUATION MODELS

GEORGE ANDREW MATYSIAK

TEGoVA
THE EUROPEAN GROUP OF VALUERS' ASSOCIATIONS

Introduction

- Second report for TEGoVA

  'Assessing the Accuracy of Individual Property Values Estimated by Automated Valuation Models'

- Objective was to identify criteria which could provide an appraiser with information on the attendant uncertainty. associated with *individual property* AVM generated valuations.
- Statistics are necessary to help *interpret* the results (the valuation) of an AVM.
- Ten items of information were recommended for inclusion in the AVM report.
- Refer to the report's *Appendix* for statistical definitions.

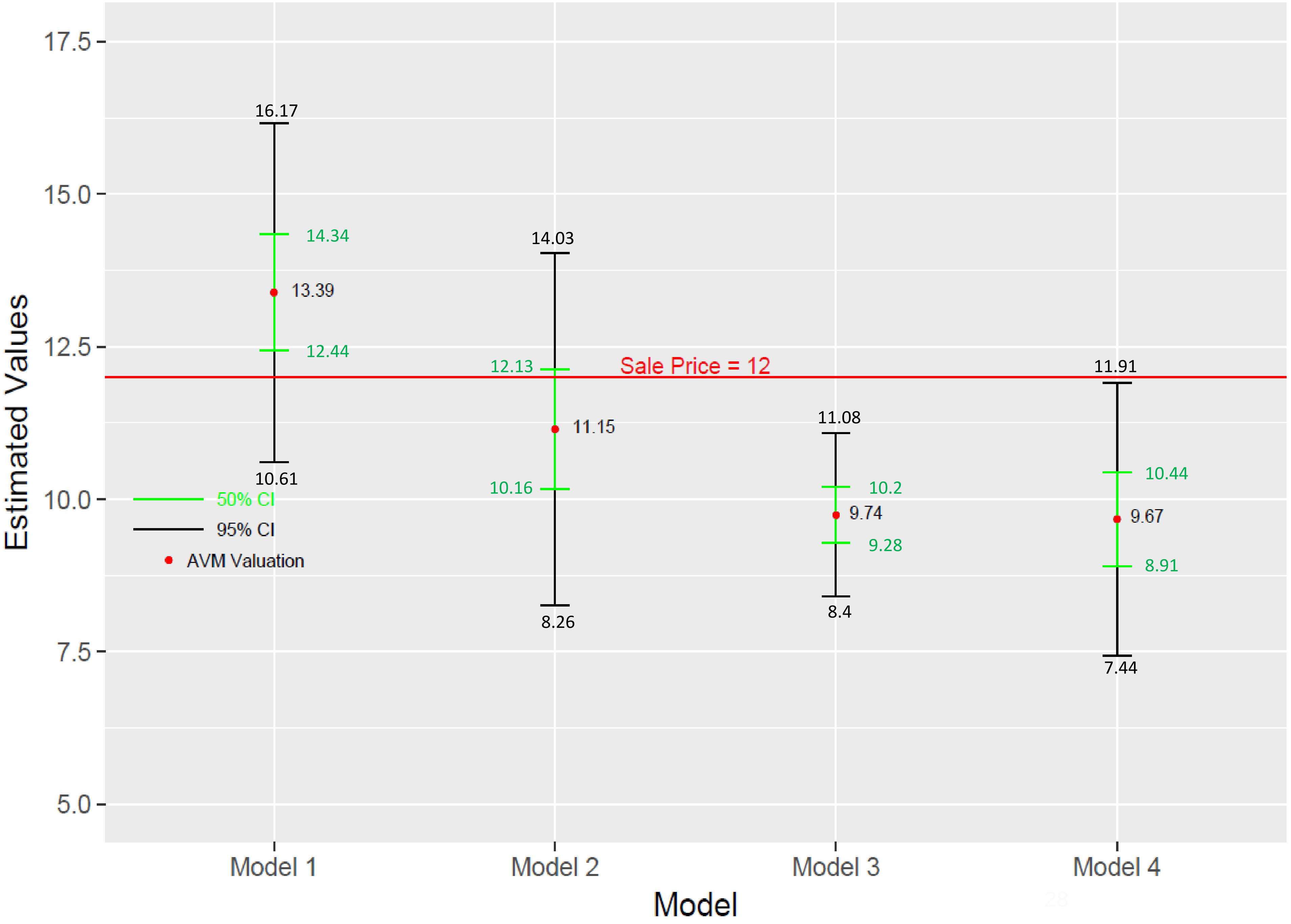
Evaluating AVM valuations

- The statistical measures are based on AVM ‘error’.
- Error measures the difference between the sale price and the AVM valuation
    - error = AVM valuation – Sales price
- Various accuracy measures using the ‘error’ can be calculated.
- In Europe, appraiser valuations are typically used in place of sale price.

TEGoVA Recommendation 1

- Confidence Intervals for the AVM valuation at:
  - i) the 50% level and
  - ii) the 95% level

Distribution of AVM Valuations for 4 Models together with 50% & 95% Confidence Intervals

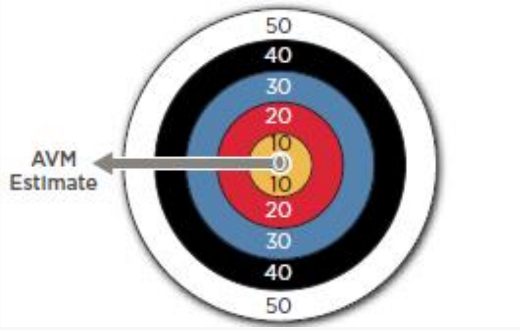


Forecast Standard Deviation (FSD)

- FSD is a statistical measure which provides the probability that the sale price falls within a range of the estimated AVM value
- The lower the FSD the smaller the range in within which sale price may lie
  - Example: If the FSD is 10%, there is a 68% (2/3rds) probability that the sale price will fall within +/-10% of the AVM estimate
  - If the AVM returns an estimate of €100,000, there is a 68% chance that the sale price will lie in the range €90,000 and €110,000

The FSDs for the four models are:

Model	FSD (%)
Model 1	9.5
Model 2	10.9
Model 3	8.8
Model 4	6.5



Target analogy for FSD Scale

The FSD tells you, with 68% statistical certainty, into which ring the sale price is likely to fall.

Appendix:
Estimating a housing model (AVM)
by regression analysis

Why model residential prices?

- Why not use "experts" to make a judgemental assessment of likely sales price?

- Judgemental assessment brings a different set of problems:

  e.g., psychologists have found that expert judgements are prone to the following biases:
    - over-confidence
    - inconsistency
    - anchoring
    - illusory patterns
    - "group-think"

- An alternative approach:

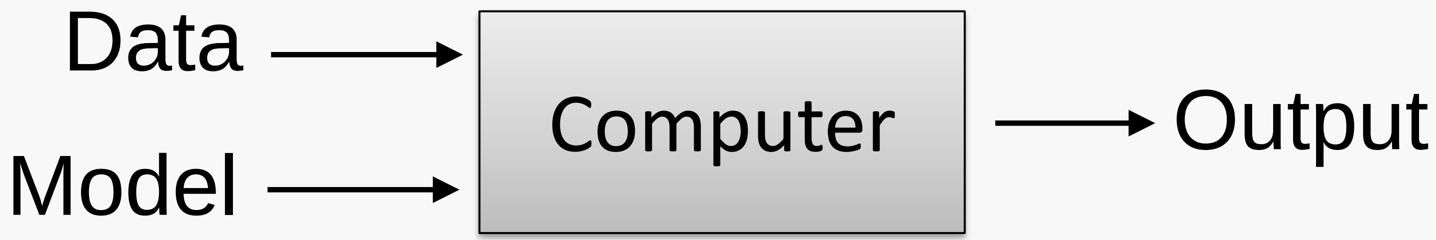
  Use a statistical/machine learning model built on solid foundations, supplemented by expert judgements and interpretation

Preamble: what is machine learning?

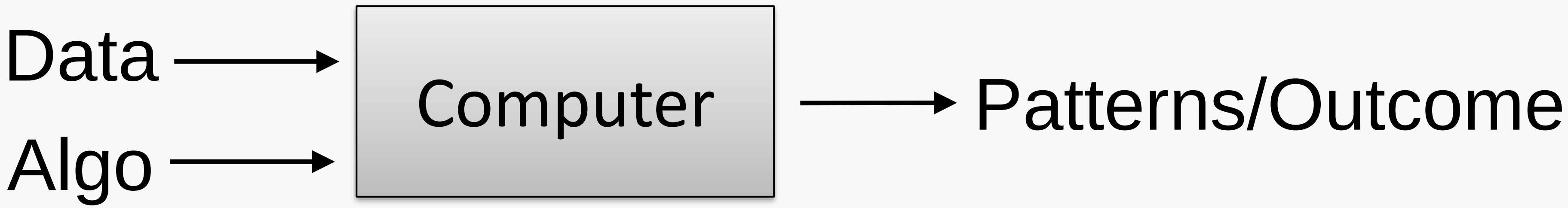
- It is very hard to write programs that solve problems like recognizing a face, for example.
  - We don’t know what type of program to write because we don’t know how our brain does this!
  - Even if we had a good idea about how to do it, the program might be horrendously complicated.
- Instead of manually writing a program, we collect lots of examples (data) that specify the correct output for a given input.
- A machine learning algorithm then takes these examples and produces a program that does the job.
- If we do it right, the program works for new cases as well as the ones we trained it on (the original data).

Preamble: machine learning (the hot approach)

Traditional programming



Machine learning (algorithms applied to data - let the data do the work!)



Rather than constructing a mathematical representation of price formation from first principles, the model, data mining involves applying algorithms and 'searching' for patterns first *and* the resulting output.

Note: https://robotwealth.com/machine-learning-financial-prediction-david-aronson

Commercial AVMs estimation: examples of some common data analysis/machine learning techniques

- Multiple regression & variants (widely used in practice)
- GIS, spatial regression, geo-location & variants
- Decision trees/random forests
- Support vector method (SVM)
- Neural networks (ANNs)
- Convolutional Neural Nets (CNNs; image recognition hot topic!)
- Many, many others
- Bottom line: exclusive and *blended* vendor models!
- Commercially sensitive!

What is Regression analysis

- Regression analysis is probably the single most widely used technique in applied data analysis and is used extensively in econometric applications and finance

  But what is regression analysis?

- It is concerned with describing and evaluating the relationship between a given variable (usually called the dependent variable) and one or more other variables (usually known as the independent variable(s))
- It is a statistical technique that attempts to describe and evaluate the relationship between variables. The relationship between one variable, the *dependent variable, and* other variables, called the *independent* (or *explanatory*) variables, is estimated
- Regression analysis is used to:
  - Predict the value of a dependent variable based on the value of at least one independent variable
  - Explain the impact of changes in independent variables on the dependent variable

Linear Regression (a naïve) example

- A real estate agent wishes to value a property with an area of 2000 square metres
- He has a sample of past sales prices and area and wants to examine the relationship between the selling price of a home and its size
- The sample consists of 10 houses:
  - Dependent variable (Y) = house price in €1000s
  - Independent variable (X) = square metres

House price and area data

House Price in €1000s (y)	Square Feet (x)
245	1400
312	1600
279	1700
308	1875
199	1100
219	1550
405	2350
324	2450
319	1425
255	1700

Is there a relationship between price and the area of houses?

A basic Linear Regression Model

[The population regression model]

Dependent Variable

Population Y intercept

Population Slope Coefficient

Independent Variable

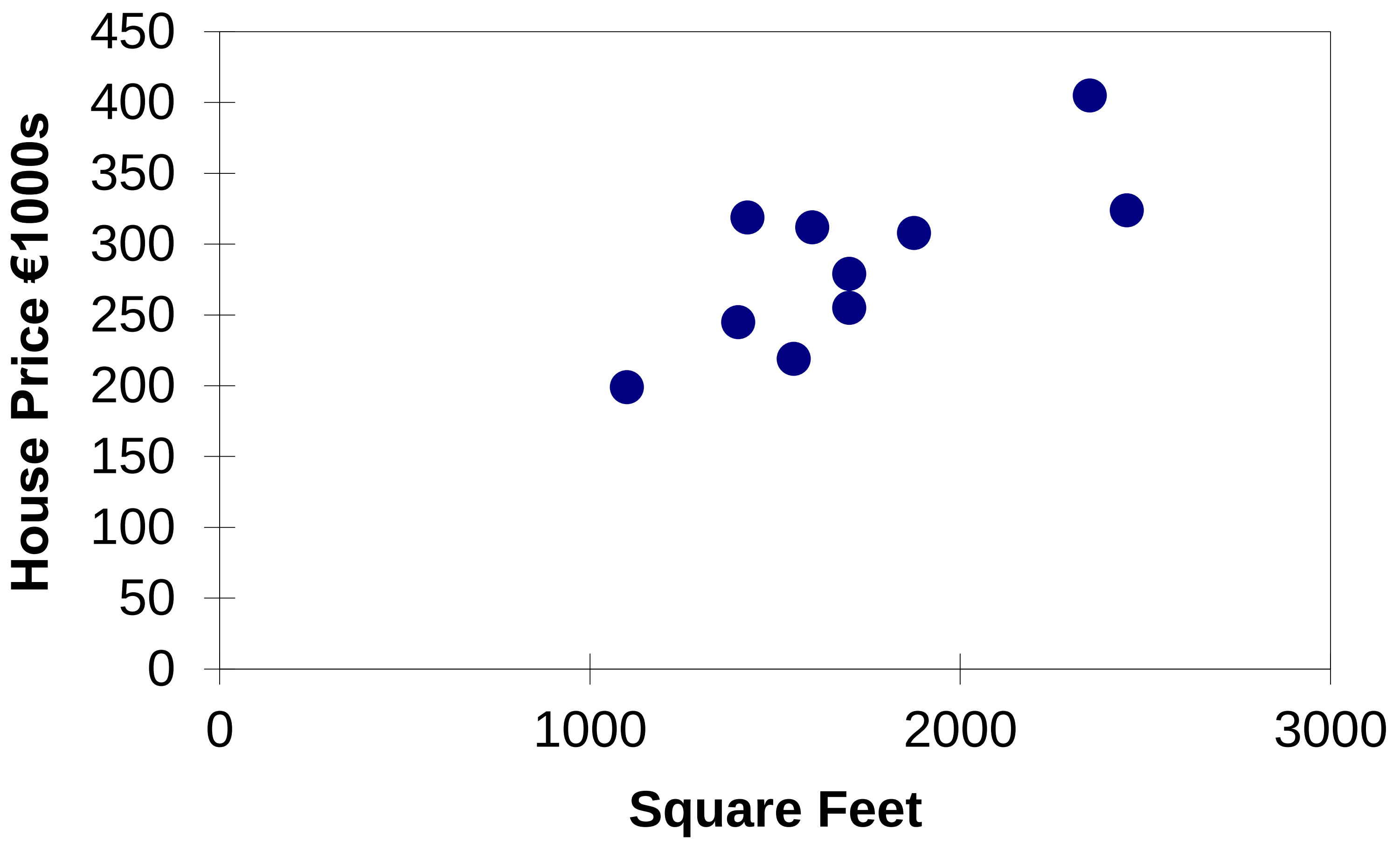
Random Error term

$$Y_i = \underbrace{\beta_0 + \beta_1 X_i}_{\text{Linear component}} + \underbrace{\varepsilon_i}_{\text{Random error component}}$$

Note: In practice we only have sample data, not the entire population

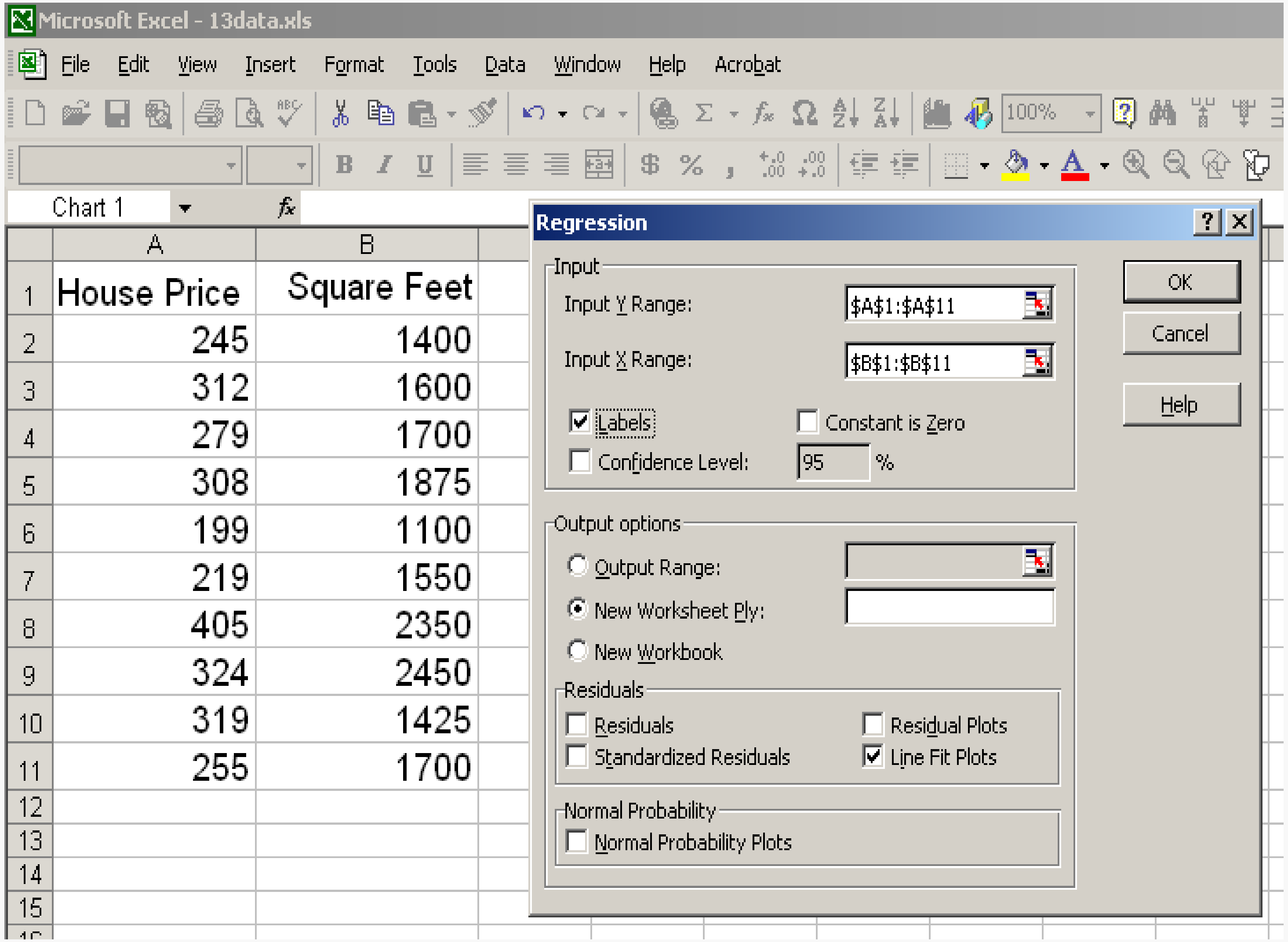
House prices and areas scatter plot

A scatter plot (or scatter diagram) can be used to show a possible relationship between two numerical variables



Linear Regression example: using Excel

Tools
>
Data Analysis
>
Regression



Linear Regression Example
Coefficient of Determination, R^2

Regression Statistics	
Multiple R	0.76211
R Square	0.58082
Adjusted R Square	0.52842
Standard Error	41.33032
Observations	10

$$R^2 = \frac{SSR}{SST} = \frac{18934.9348}{32600.5000} = 0.58082$$

58.08% of the variation in house prices is explained by variation in square feet

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	18934.9348	18934.9348	11.0848	0.01039
Residual	8	13665.5652	1708.1957		
Total	9	32600.5000			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	98.24833	58.03348	1.69296	0.12892	-35.57720	232.07386
Square Feet	0.10977	0.03297	3.32938	0.01039	0.03374	0.18580

Confidence Interval estimate for the slope ("Beta coefficient")

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	**98.24833**	**58.03348**	**1.69296**	**0.12892**	**-35.57720**	**232.07386**
Square feet	**0.10977**	**0.03297**	**3.32938**	**0.01039**	**0.03374**	**0.18580**

Conclusion: There is a significant relationship between house price and square feet at the .05 level of significance

Linear Regression Excel output

Regression Statistics	
Multiple R	0.76211
R Square	0.58082
Adjusted R Square	0.52842
Standard Error	41.33032
Observations	10

The regression equation is:

$$\text{house price} = 98.24833 + 0.10977\,(\text{square metres})$$

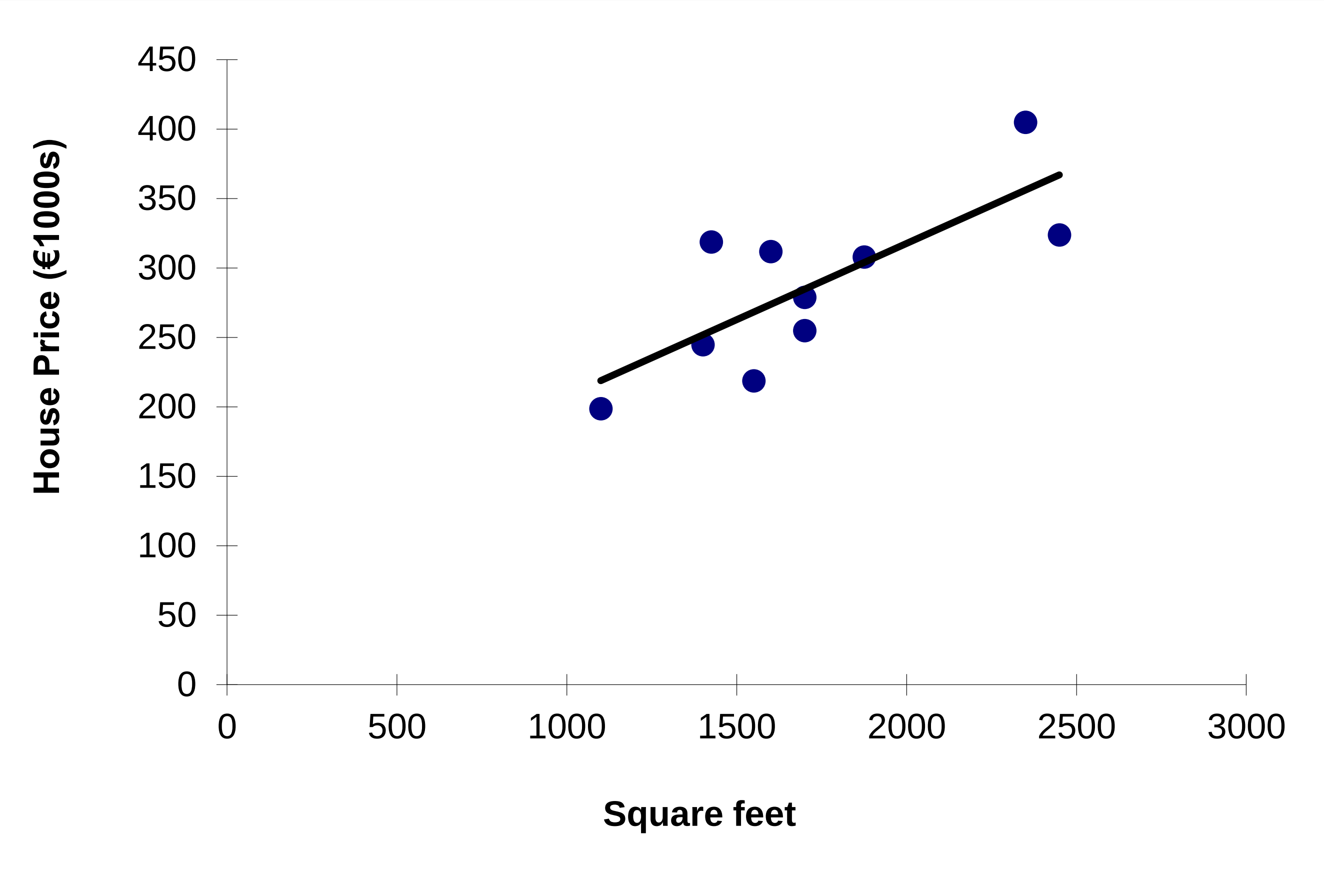
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	18934.9348	18934.9348	11.0848	0.01039
Residual	8	13665.5652	1708.1957		
Total	9	32600.5000			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	98.24833	58.03348	1.69296	0.12892	-35.57720	232.07386
Square feet	0.10977	0.03297	3.32938	0.01039	0.03374	0.18580

Linear Regression example graphical representation

House price model: scatter plot and regression line



Slope
= 0.10977

Intercept
= 98.248

$$\text{house price} = 98.24833 + 0.10977\ (\text{square feet})$$

Linear Regression making predictions

Predict the price for a house with 2000 square feet:

$$
\begin{aligned}
&\text{house price} \\
&= 98.25 \ + 0.1098\ (\text{sq.ft.}) \\
&= 98.25 \ + \ 0.1098(2000) \\
&= 317.85
\end{aligned}
$$

The predicted price for a house with 2000 square feet is 317.85(€1,000s) = €317,850

Standard error: distribution of the model's errors

Regression Statistics	
Multiple R	**0.76211**
R Square	**0.58082**
Adjusted R Square	**0.52842**
Standard Error	**41.33032**
Observations	**10**

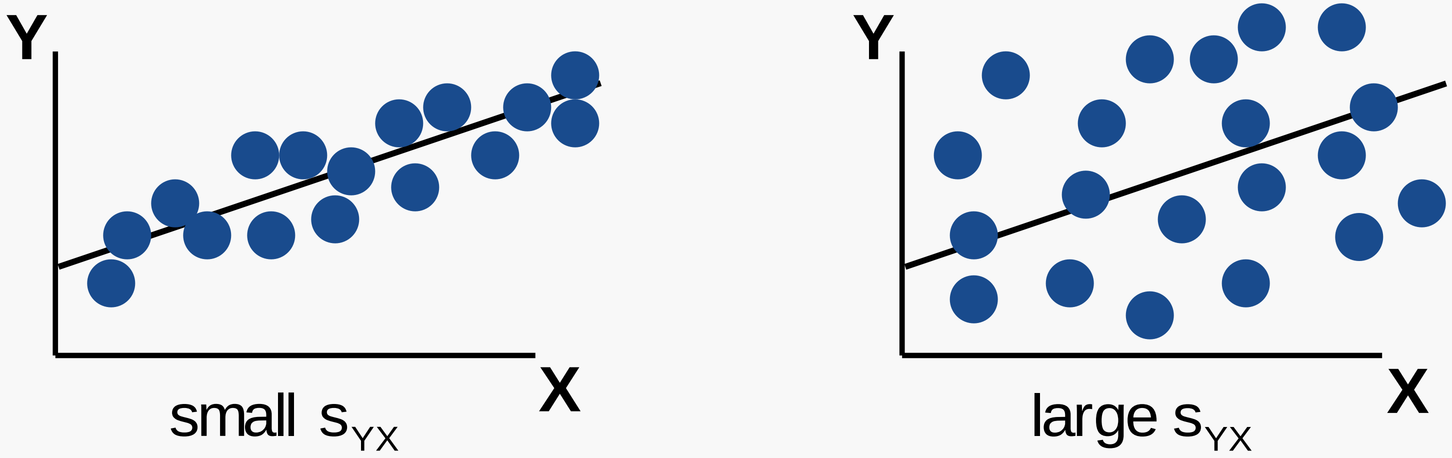
$$S_{YX} = 41.33032$$

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	**1**	**18934.9348**	**18934.9348**	**11.0848**	**0.01039**
Residual	**8**	**13665.5652**	**1708.1957**		
Total	**9**	**32600.5000**			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	**98.24833**	**58.03348**	**1.69296**	**0.12892**	**-35.57720**	**232.07386**
Square Feet	**0.10977**	**0.03297**	**3.32938**	**0.01039**	**0.03374**	**0.18580**

Comparing Standard Errors

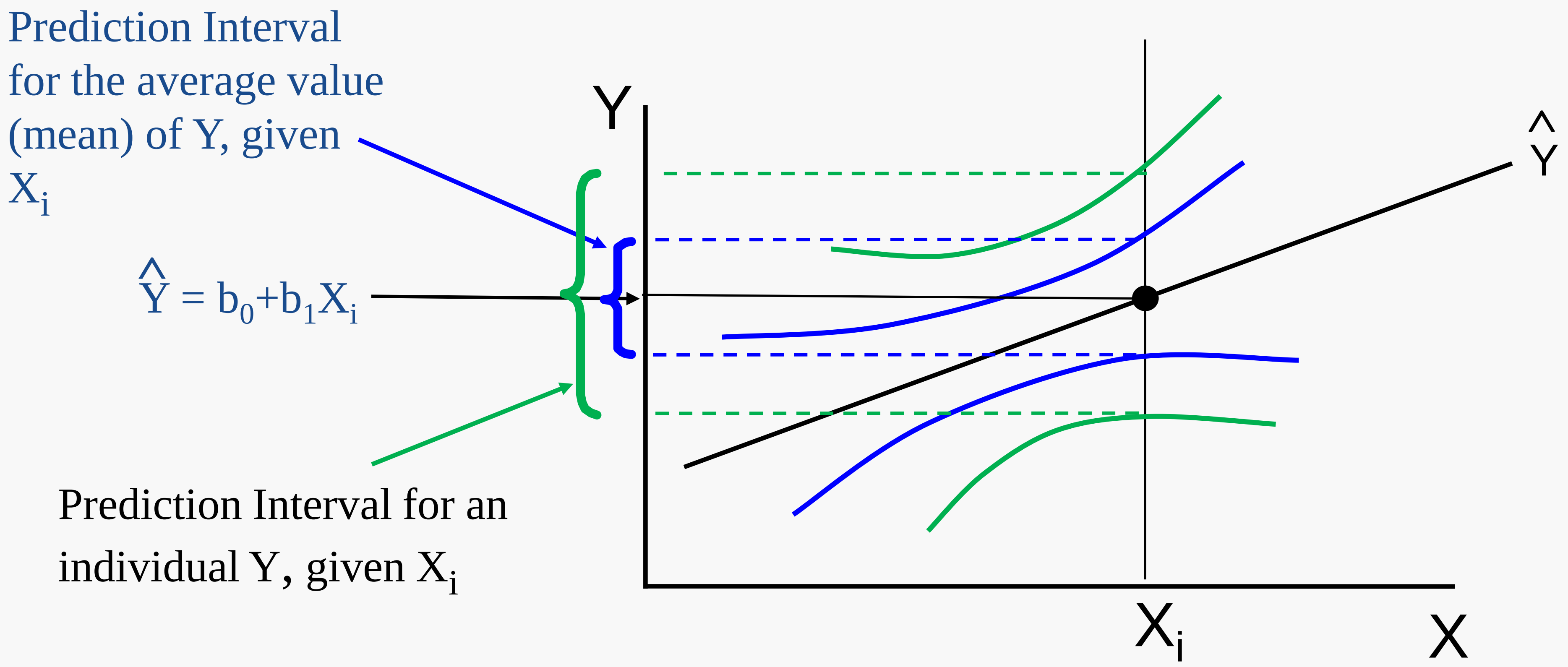
S_{YX} is a measure of the variation of observed Y values from the regression line



The magnitude of S_{YX} should always be judged relative to the size of the Y values in the sample data

“Mean” value and individual value uncertainty: Prediction Intervals

Goal: Form intervals around Y to express uncertainty about the value of Y_i for a given value of X_i



Prediction interval for the mean (average) of house price values

Prediction Interval Estimate for $\mu_{Y|X=X_i}$

Find the 95% confidence interval for the mean price of 2,000 square-foot houses

Predicted Price $\hat{Y}_i$ = € 317.85 (€1,000s)

$$\hat{Y} \pm t_{n-2} S_{YX} \sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2}} = 317.85 \pm 37.12$$

The confidence interval endpoints are €280.66 and €354.90, or from €280,660 to €354,900

Prediction interval for an *individual* house price

Prediction Interval Estimate for $Y_{X=X_i}$

Find the 95% prediction interval for an *individual* house with 2,000 square feet

Predicted Price $\hat{Y}_i$ = €317.85 (€ 1000s)

$$\hat{Y} \pm t_{n-1} S_{YX} \sqrt{1 + \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2}} = €317.85 \pm €102.28$$

The prediction interval endpoints are €15.50 and €420.07, or from €215,500 to €420,070

Estimating Multivariate Regression Models

- It's unlikely that many dependent variables can be fully explained by a single independent variable
- Consequently, it's necessary to employ *multivariate regression models*
- The multivariate linear regression model with K independent variables is formulated as:

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \ldots + b_K X_{Ki} + e_i$$

Example: a more realistic relationship for house price determination?

- A multivariate regression of a relationship may be expressed by the equation:

$$P_i = X_{1i} + X_{2i} + X_{3i} + X_{4i} + \varepsilon_i$$

dependent variable: P_i = Price of property

independent variables: X_{1i} = Area, X_{2i} = Number of bedrooms

X_{3i} = Age, X_{4i} = Location

ε_i = random error term

The Multiple Regression Model

Idea: Examine the linear relationship between 1 dependent (Y) & 2 or more independent variables (X_i)

Multiple Regression Model with k Independent Variables:

$$Y_i = \underbrace{\beta_0}_{\text{Y-intercept}} + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \underbrace{\varepsilon_i}_{\text{Random Error}}$$

Population slopes

Example: King County, Seattle

Model 8: OLS, using observations 1-5000
Dependent variable: price

	coefficient	std. error	t-ratio	p-value	
const	−34154.2	1202.85	−28.39	1.52e-164	***
waterfront	1080.63	36.7729	29.39	3.19e-175	***
sqft_above	0.110568	0.00722140	15.31	9.51e-052	***
sqft_living15	0.0821189	0.00835936	9.824	1.43e-022	***
lat	705.411	25.3790	27.80	3.20e-158	***
grade	102.480	5.01480	20.44	3.30e-089	***

Mean dependent var	538.9748	S.D. dependent var	386.1188
Sum squared resid	3.04e+08	S.E. of regression	246.5269
R-squared	0.592758	Adjusted R-squared	0.592351
F(5, 4994)	1453.798	P-value(F)	0.000000
Log-likelihood	−34629.05	Akaike criterion	69270.09
Schwarz criterion	69309.19	Hannan-Quinn	69283.80

Note: Data source ‘eRealProperty’ catalogue, obtained from Kaggle, covering the period May 2014 to May 2015.
Model estimation George Matysiak.

Example: King County, Seattle

Grades: OLS, using observations 1-5000
Dependent variable: price

	coefficient	std. error	t-ratio	p-value	
sqft_above	0.0504820	0.00744965	6.776	1.37e-011	***
sqft_living15	0.0962492	0.00862126	11.16	1.33e-028	***
Dgrade_3	18.1517	177.348	0.1024	0.9185	
Dgrade_4	24.5046	103.221	0.2374	0.8124	
Dgrade_5	53.3860	34.7287	1.537	0.1243	
Dgrade_6	105.152	16.0297	6.560	5.93e-011	***
Dgrade_7	172.087	15.0730	11.42	8.05e-030	***
Dgrade_8	250.435	19.1797	13.06	2.44e-038	***
Dgrade_9	399.133	25.3454	15.75	1.44e-054	***
Dgrade_10	618.225	31.8098	19.44	3.67e-081	***
Dgrade_11	882.126	42.9071	20.56	3.21e-090	***
Dgrade_12	1710.49	67.0450	25.51	4.75e-135	***
Dgrade_13	3414.45	182.212	18.74	9.06e-076	***
waterfront	954.234	37.7783	25.26	1.44e-132	***

Mean dependent var	538.9748	S.D. dependent var	386.1188
Sum squared resid	3.13e+08	S.E. of regression	250.4173
R-squared	0.580477	Adjusted R-squared	0.579383
F(13, 4986)	530.6865	P-value(F)	0.000000
Log-likelihood	−34703.33	Akaike criterion	69434.65
Schwarz criterion	69525.89	Hannan-Quinn	69466.63

Note: Data source 'eRealProperty' catalogue, obtained from Kaggle, covering the period May 2014 to May 2015. Model estimation George Matysiak.

Stages in the modelling/predicting residential prices

- Define the prediction problem
- Look at relevant theory & techniques
- Collect and analyse the data
- Define/specify and estimate the model
  - test robustness/diagnostics, forecast and evaluate
- Distribute the predicted sales value
- Provide a prediction interval (range) and/other uncertainty measure (see *TEGoVA* report)
- Monitor and evaluate the predictions when actual sales values become known. Where did it go wrong?

Evaluating the 'quality' of a regression equation

- There may be a tendency to accept regression results without thinking about their meaning or validity
- The modeler should carefully think about the evaluation of every aspect of an equation
- This includes:
    1. Underlying theory
    2. Quality of the data
    3. Testing the regression and the underlying assumptions thoroughly
    4. Do the results make sense?

Evaluating the ‘quality’ of a regression equation

5. How well do the estimated coefficients correspond to the expectations before the data were collected?

6. Are all the obviously important variables included in the equation?

7. Has the most theoretically logical functional form been used?

8. Does the regression appear to be free of major ‘econometric’ problems?

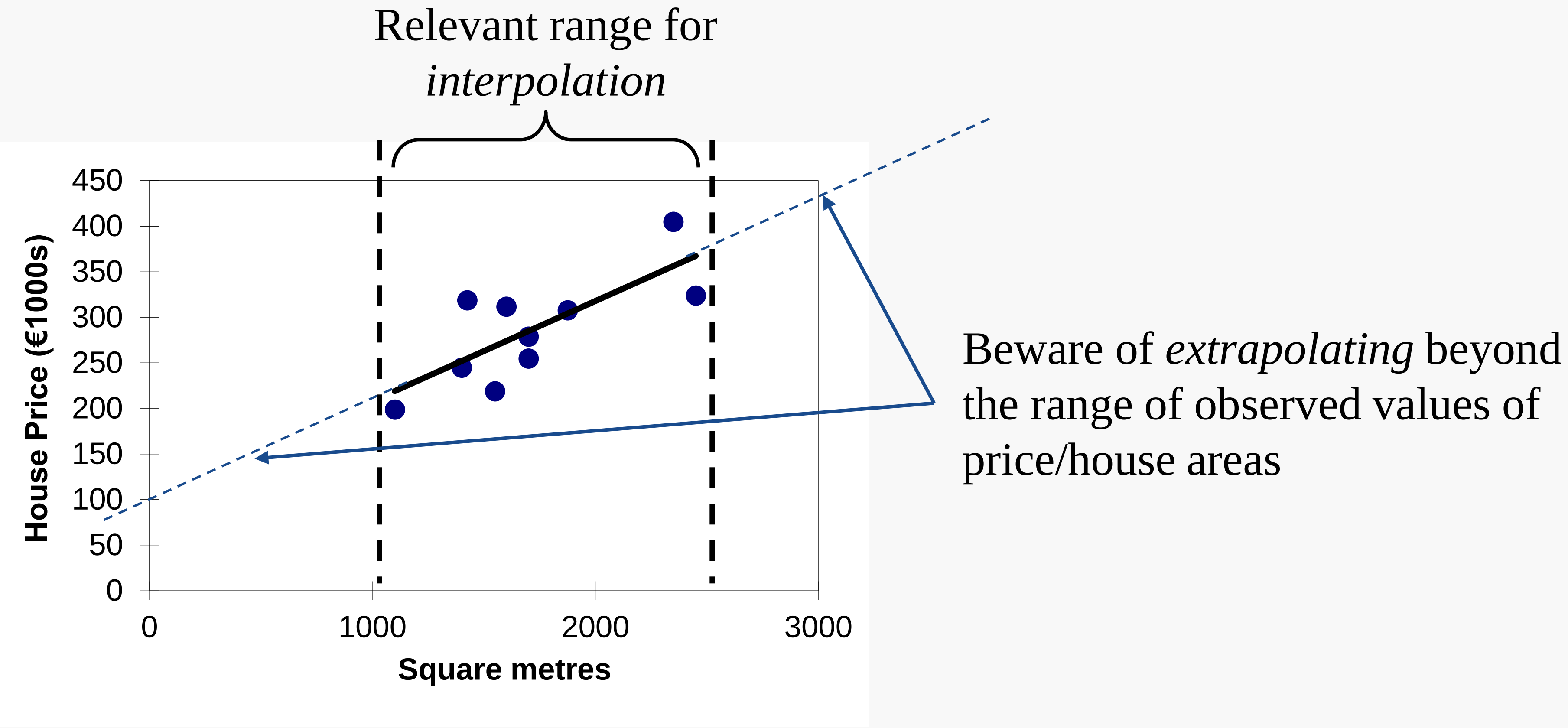
9: Data mining issues

How can model predictions go wrong?

- Are the input variables reliable?
- Market shock/surprises
- Model errors: the model is not perfect
- Adjustments to pure model generated predictions
- Structural breakdown/model failure – the world changes!
- Predictions are outside the range of experience
- Random component could be ‘large’ in some situations
- Thin markets i.e. very little or very specific evidence is available
  - ➢ is the sample evidence (comparables) representative?

…finally, be aware of how you use the estimated regression to make predictions

- When using a regression model for prediction of values, predict within the *relevant range* of data used in estimating the model



ASAVAL Conference

Valuation for the Financial Sector:

Challenges in the short and medium term

19th November, 2021

AVMs – Opportunity or Threat?

A holistic view of technology at the service of the evaluator

Professor George Matysiak